ABSTRACT OF THE DISCLOSURE

This invention provides a method, a system and a computer program for recognizing technical terms. In the preferred embodiment the technical terms are chemical names, and in a most preferred embodiment the technical terms are organic chemical names. A computer program product stores in a computer readable form a set of computer program instructions for directing at least one computer to process a text document. The set of computer program instructions include instructions for assigning corresponding associated parts of speech to words found in the document. The instructions for assigning include instructions to apply a plurality of regular expressions, rules and a plurality of dictionaries to recognize organic chemical name fragments, to combine recognized organic chemical name fragments into a complete organic chemical name, and to assign the complete organic chemical name with one part of speech. The regular expressions include a plurality of patterns, individual ones of which are comprised of at least one of characters, numbers and punctuation. For example, the punctuation can comprise at least one of parenthesis, square bracket, hyphen, colon and semi-colon, and the characters can comprise at least one of upper case C, O, R, N and H, and further comprise strings of at least one of lower case xy, ene, ine, yl, ane and oic.